



IDRC-TS14f

**Un système interactif
sur mini-ordinateur pour
la recherche documentaire
et la gestion de
bibliothèques**

Faye A. Daneliuk

Le Centre de recherches pour le développement international, société publique créée en 1970 par une loi du Parlement canadien, a pour mission d'appuyer des recherches visant à adapter la science et la technologie aux besoins des pays en voie de développement; il concentre son activité dans cinq secteurs: agriculture, alimentation et nutrition; information; santé; sciences sociales; et communications. Le CRDI est financé entièrement par le Gouvernement du Canada, mais c'est un Conseil des gouverneurs international qui en détermine l'orientation et les politiques. Établi à Ottawa (Canada), il a des bureaux régionaux en Afrique, en Asie, en Amérique latine et au Moyen-Orient.

©Centre de recherches pour le développement international, 1979
Adresse postale: B.P. 8500, Ottawa, Canada K1G 3H9
Siège: 60, rue Queen, Ottawa

Danieliuk, F. A.
CRDI, Ottawa CA

IDRC-TS14f

Un système interactif sur mini-ordinateur pour la recherche documentaire et la gestion de bibliothèques. Ottawa, Ont., CRDI, 1979. 19p.

/Publication CRDI/. Monographie sur le /système d'information/ à base de mini/ordinateur/ au /CRDI/- discute l'/étude de viabilité/ de la mise en œuvre du système d'information bibliographique sur un miniordinateur, la /conception de système/ et le travail lié au /programme d'ordinateur/ effectué au /CRDI/; inclut une bibliographie.

CDU: 002

ISBN: 0-88936-192-4

Édition microfiche sur demande

Un système interactif sur mini-ordinateur pour la recherche documentaire et la gestion de bibliothèques

Faye A. Daneliuk

*Administrateur du projet sur le mini-ordinateur,
Centre de recherches pour le développement international*

La présente monographie est la version révisée d'un texte de l'auteur paru dans The application of inexpensive minicomputers to information work, AGARD Lecture Series no. 92, ISBN: 92-835-1276-6 (Organisation du Traité de l'Atlantique du Nord, Neuilly-sur-Seine, mars 1978).

Avant-propos

Au cours des dix dernières années, la vitesse et la fiabilité d'exploitation des ordinateurs se sont accrues de façon notable. Au cours de la même période, grâce aux circuits mini-électroniques intégrés, la taille et le coût des mémoires ont été réduits selon un facteur allant de 100 à 1000. Maintes fonctions complexes du matériel que seuls pouvaient offrir, il y a quelques années, des ordinateurs coûteux de grande puissance, sont maintenant standard dans des mini-ordinateurs peu coûteux. Les fabricants de ceux-ci offrent également un logiciel comportant un système d'exploitation assez évolué et répondant aux besoins variés des utilisateurs.

C'est la raison pour laquelle les utilisateurs, lorsqu'ils doivent procéder à l'automatisation de nouvelles applications, se tournent de plus en plus vers le mini-ordinateur. Des gouvernements, des institutions et des industries disposant de budgets limités ont opté pour des mini-ordinateurs afin de satisfaire à leurs propres besoins. Certains de ces utilisateurs ont non seulement fait la preuve que des systèmes fondés sur des mini-ordinateurs sont rentables, mais font un usage particulièrement complexe de ces appareils.

La présente brochure contient une description du système de traitement des données bibliographiques enregistrées sur le mini-ordinateur du Centre de recherches pour le développement international, qui a été mis au point à Ottawa au cours des deux années et demie écoulées. Ce système est maintenant opérationnel grâce à notre mini-ordinateur maison Hewlett-Packard 3000 Série II et offre un grand nombre de procédures automatiques permettant de répondre aux besoins de la bibliothèque du Centre et d'extraire des données de plusieurs grandes bases. Lors de la conception de ce système, on s'est largement inspiré de l'expérience acquise avec ISIS (Ensemble de systèmes intégrés d'information), programme-produit élaboré à Genève par le Bureau international du travail et destiné à être traité sur les ordinateurs IBM 360/370. Bien que le système sur mini-ordinateur du CRDI soit compatible avec ISIS, il n'en constitue en rien un remaniement.

Pour élaborer le système du CRDI, on s'est inspiré des connaissances les plus récentes en structure de base de données et de technologie informatique; le système, suffisamment généralisé, permet la création de bases de données pour divers types d'applications.

L'une des principales raisons ayant décidé le CRDI à se lancer dans ce projet a été son appui très actif à l'égard des systèmes d'information internationaux. Beaucoup de pays développés ou en voie de l'être ont manifesté un vif intérêt quant à l'acquisition éventuelle d'un ordinateur peu coûteux qui leur donne accès aux systèmes d'information internationaux décentralisés. Le système sur mini-ordinateur du CRDI tient compte des exigences de systèmes décentralisés, tels AGRIS (Système international d'information sur les sciences et la technologie agricoles) et DEVSIS (Système d'information sur les sciences du développement), notamment du traitement de bases de données structurées ou non selon les règles de l'UNISIST (Programme intergouvernemental de coopération en matière d'information scientifique et technologique) et de la diffusion sélective de l'information (SDI).

On assiste en ce moment à des efforts concertés en vue de la production d'un programme-produit complet (incluant l'ensemble de la documentation, et les instruments de formation et d'exécution) et on estime qu'il pourra être diffusé dans un proche avenir.

Sultan Kassum

Administrateur, Systèmes de
l'informatique
Division des sciences de
l'information
Centre de recherches pour le
développement international*

**A quitté le CRDI en juin 1978.*

Introduction et historique

Le Centre de recherches pour le développement international est une société d'intérêt public, créée en 1970 en vertu d'une loi du Parlement canadien. Il a été constitué en organisme autonome, dirigé par un Conseil des gouverneurs de 21 membres (provenant de différents pays), qui fixe les grandes lignes de sa politique et en approuve chacun des projets. Le Centre a son siège à Ottawa et des bureaux régionaux à Singapour, au Caire, à Bogota et à Dakar.

Aux termes de la Loi, le Centre a pour objectifs "d'entreprendre, d'encourager, de soutenir et de poursuivre des recherches sur les problèmes des régions du monde en voie de développement et sur les moyens d'application et d'adaptation des connaissances scientifiques, techniques et autres au progrès économique social de ces régions" et "d'aider les régions en voie de développement à se livrer à la recherche scientifique, à acquérir les techniques innovatrices et les institutions requises pour résoudre leurs problèmes". A cette fin, il a reçu les pouvoirs de "créer, maintenir et exploiter des centres de renseignements et d'information et des installations en vue de la recherche ou d'autres activités connexes à ses objets" et d'"entreprendre et poursuivre la recherche et le développement technique, y compris l'établissement et le fonctionnement de toute installation ou projet pilote, jusqu'au point où les résultats appropriés de ces recherches et de ce développement peuvent être appliqués".

Le Centre comprend les 5 divisions suivantes correspondant à autant de programmes: Sciences de la santé; Sciences de l'agriculture, de l'alimentation et de la nutrition; Sciences sociales et Ressources humaines; Sciences de l'information; et, enfin, Communications. La Division des sciences de l'information est le seul exemple d'un programme créé par un organisme d'aide et ayant pour but précis d'appuyer des projets de diffusion de l'information dans les pays en voie de développement. C'est cette division qui s'est lancée dans l'application de l'informatique au travail de l'information.

En 1972, le Conseil des Gouverneurs du Centre a approuvé un projet portant sur des systèmes d'information informatisés. Le projet poursuivait les quatre buts suivants:

(1) faire l'acquisition d'un système en direct permettant au Centre d'informatiser les opérations de sa bibliothèque;

2) créer une base de données assimilable par l'ordinateur contenant toute la documentation du Centre relative au développement;

3) collaborer à l'échelle internationale avec d'autres organismes désirant travailler à la mise sur pied d'un réseau coopératif utilisant un système commun et

4) acquérir l'expérience permettant au personnel du CRDI de contribuer à l'établissement de centres d'entrée/sortie dans les régions en voie de développement.

Dans le cadre de ce projet, le Centre a fait l'acquisition du système ISIS élaboré en quelques années par le Bureau international du travail à Genève. ISIS a été retenu parmi un certain nombre d'autres systèmes, y compris des systèmes commerciaux, pour les raisons suivantes: il permettait l'introduction et l'extraction de données en mode interactif; il offrait un grand nombre de fonctions de traitement en différé en vue de la gestion de bibliothèque; il était international. En effet, à cette époque, l'Office mexicain pour l'information et le travail était en train de l'adopter, la SAFAD (Agence suédoise pour le développement en administration) l'utilisait, le BIT et l'ONUDI (Organisation des Nations Unies pour le développement industriel) l'employaient à Genève, et le Gouvernement roumain l'avait déjà mis en place au Centre d'administration pour les documentalistes et bibliothécaires de Bucarest; il permettait le transfert de bases de données par l'intermédiaire de bandes magnétiques correspondant à la norme ISO 2709 et, enfin, il donnait immédiatement accès à deux bases de données contenant de la documentation sur le développement.

Programmé en langage Assembleur IBM 360, ISIS exigeait une installation informatique d'une puissance suffisante pour employer un ordinateur 360 et le système d'exploitation sur disque DOS. Le CRDI a donc dû mettre en place ISIS auprès d'un centre de traitement à façon, l'acquisition d'un ordinateur 360 et des appareils de transmission connexe représentant un investissement trop important. Le recours à un tel service à bureau n'est pas nécessairement bon marché et c'est la raison pour laquelle la formulation du projet laissait la Division libre à l'avenir de louer ou d'acheter un ordinateur (compatible avec le logiciel ISIS, ou pour lequel il faudrait créer un nouveau logiciel). En 1975, après deux années et demie d'exploitation d'ISIS, il a été décidé d'étudier la possibilité d'acquérir un ordinateur qui serait installé sur place.

Acquisition d'un ordinateur

En avril 1975, M. Gordon Somerfield a été engagé à titre de conseiller par la Division des sciences de l'information, afin d'étudier la faisabilité de l'installation d'un système de données bibliographiques sur mini-ordinateur. Tout en tenant compte de facteurs tels que les coûts et l'élaboration d'un logiciel, le rapport concluait par l'affirmative, ce qui justifiait un examen plus approfondi de notre part.

Au cours des 8 mois qui ont suivi, on a procédé à une évaluation critique de tous les fabricants stables de mini-ordinateurs et de leurs produits. Bien que le coût du matériel soit important, ce n'était pas là le seul facteur qui ait été pris en considération. Étant donné que le Centre entendait élaborer son propre logiciel, l'étendue et la fiabilité de celui du fabricant ont été minutieusement examinées. Notre produit final (un logiciel d'information) n'était pas destiné seulement au CRDI mais également à des régions qui utiliseraient une installation peu coûteuse mais fiable, plutôt qu'un ordinateur de grande puissance. Il fallait donc que nous soyons assurés au moins dans une certaine mesure que le fabricant resterait en affaires pendant un certain temps. Le fabricant devait également être en mesure d'assurer un certain service en Afrique, dans le Sud-Est asiatique et en Amérique Latine, puisque nous espérions mettre notre logiciel à la disposition d'institutions de ces régions. (Fait à noter, aucun des fabricants ne pouvait satisfaire à cette dernière exigence!)

Au cours de la même période, nous avons contacté d'autres institutions qui procédaient à l'élaboration de systèmes d'information pour mini-ordinateurs. Certaines d'entre elles avaient porté leur choix sur un ordinateur apte à traiter un système spécialisé et nous avons d'abord été enclins à adopter la même solution. En effet, on comprend facilement que si un ordinateur est destiné à une seule application, le logiciel du fabricant n'a pas à être extrêmement évolué. On peut toujours se résigner à rédiger des programmes de traitement des entrées/sorties sur terminal, des systèmes de gestion de fichiers ou tout programme dont l'emploi reste restreint. Au cours de cette période d'évaluation, cependant, il y a eu un dialogue constant entre la Division des sciences de l'information et les autres divisions du Centre. On s'est alors rendu compte que l'acquisition d'un ordinateur un peu plus complexe pourrait grandement bénéficier au Centre lui-même. Cette décision restreignait le nombre des fournisseurs possibles.

Au début de l'année 1976, la Division des sciences de l'information a présenté au Conseil des gouverneurs un projet comportant à la fois l'acquisition d'un système de mini-ordinateur et l'élaboration d'un logiciel. Approuvé par le Conseil, ce projet faisait valoir les trois principales raisons suivantes en faveur d'un mini-ordinateur:

- 1) réduire les frais d'exploitation d'ISIS pour le CRDI — facteur déterminant, car les coûts d'exploitation auprès d'un centre de traitement à façon peuvent être extrêmement élevés;

- 2) déterminer, après une analyse de coûts-rendement, la meilleure installation de mini-ordinateurs à offrir aux centres nationaux des pays en voie de développement, incluant des programmes et pouvant traiter les systèmes AGRIS, DEVSIS et ISIS; et

- 3) établir la base d'un centre canadien d'entrée/sortie en vue de DEVSIS.

Les trois membres de l'équipe informatique de la Division des sciences de l'information ont rédigé un appel d'offres où ils insistaient sur trois caractéristiques: 1) la puissance du système d'exploitation; 2) l'existence et la fiabilité du logiciel du fabricant; et 3) l'aptitude de l'ordinateur à traiter un éventail de travaux. La même équipe a mis au point et effectué des essais en laboratoire permettant de vérifier les trois caractéristiques ci-dessus. Le système 3000 de Hewlett-Packard a été retenu. Bien que le HP-3000 présente certaines lacunes, qu'on s'applique à corriger en ce moment seulement (à savoir le traitement par bandes immatriculées et un support pour unités de disques à accès réservé), il offre, du point de vue matériel, quelques caractéristiques très intéressantes (en particulier sa structure par piles) et son système d'exploitation fiable et intégré, le MPE (Multiprogramming Executive); ces qualités ont contribué au choix de l'appareil. Le matériel a été livré en août 1976 et on a commencé alors à élaborer le logiciel.

Élaboration du système

Une fois l'ordinateur choisi, le Centre a porté son intérêt sur l'élaboration d'un système. En mars 1976, il a fallu prendre une décision finale, celle de créer un nouveau logiciel pour notre système d'information. D'autres solutions s'offraient à nous. Nous aurions pu nous contenter de recoder les programmes

ISIS en vue du HP-3000 ou encore d'adopter le programme-produit de gestion de bases de données (IMAGE) mis au point par Hewlett-Packard pour le 3000. La première formule, bien qu'elle eût constitué une solution rapide et facile à un problème d'élaboration de système, n'aurait pas été réaliste à l'usage puisqu'elle n'aurait pas tiré parti des caractéristiques spéciales du HP-3000. Quant à la deuxième, elle a dû être considérée avec plus de soin mais a été finalement rejetée pour des raisons invoquées par ceux qui ont utilisé des systèmes bibliographiques. En effet, le HP-3000 n'offrait la possibilité de traiter aucune des données suivantes: 1) de véritables articles de longueur variable, 2) des zones de longueur variable et apparaissant de façon variable, 3) des sous-zones, 4) des indicatifs intercalés dans un texte et 5) de longs résumés descriptifs. L'élaboration d'un nouveau système s'imposait donc sans hésitation.

Lors de la première étape d'élaboration du système, les lignes directrices suivantes ont été adoptées:

1) généralité d'application: le système devait être d'application aussi générale que possible;

2) modularité: le système devait être entièrement modulaire, ce qui facilite la mise à jour et l'extension;

3) cloisonnement: les fonctions d'application devaient être indépendantes des fonctions de gestion des bases de données;

4) adaptation aux besoins des utilisateurs — a) le système devait être suffisamment souple pour traiter des données présentées sous les formes les plus diverses; b) il devait être suffisamment simple à comprendre pour être mis en place et utilisé sans trop d'effort; c) il devait employer un langage qui plaise suffisamment aux utilisateurs pour que ceux-ci en fassent réellement usage et d) il devait offrir tout un choix de sorties;

5) conformité avec le but recherché: le système devait — a) être en mesure d'accepter en entrée les sorties d'autres systèmes d'information; b) être adapté à long terme à un organisme de petite envergure; et c) être compatible avec d'autres systèmes internationaux, en particulier ISIS et AGRIS; et

6) rentabilité: le système de base devait être en exploitation en décembre 1977 pour que le Centre n'ait plus à recourir au service à bureau ayant assuré l'exploitation d'ISIS.

Fondements théoriques

Bien que des lignes directrices soient très importantes lors de l'élaboration d'un système, elles ne sont pas suffisantes. Ce système doit également s'inscrire dans un cadre théorique qui lui donne une cohérence, autrement difficile à atteindre. On a donc étudié avec soin les trois principales théories de gestion de bases de données, qui sont la technique CODASYL (par réseau), la technique hiérarchique et la technique relationnelle. Cette dernière a finalement été retenue parce que le modèle de données qui en découlait offrait un certain nombre d'avantages inhérents, contrairement aux autres modèles.

Ce modèle se fonde sur la théorie mathématique des relations. Bien qu'il existe une définition mathématique d'une relation (voir à ce sujet Date 1975 et

Codd 1970), on peut considérer qu'une relation est un ensemble de différents articles de base, composés d'un nombre quelconque de données élémentaires. Il s'ensuit que l'on peut représenter une relation au moyen d'un tableau à deux dimensions où 1) il n'existe pas deux lignes identiques; 2) chaque position de ligne ou de colonne de la table ne recèle qu'une seule valeur et non pas un ensemble de valeurs (en d'autres termes, les groupes répétitifs ne sont pas permis); 3) les colonnes sont homogènes et 4) les zones ne contenant pas d'indicatif dépendent sur le plan fonctionnel de l'indicatif de base. Etant donné cette structure, un ensemble généralisé d'opérateurs de relation est défini en vue de la manipulation des données, aussi bien au niveau du domaine (ou champ) qu'au niveau de la relation (ou fichier). Ces opérateurs sont définis dans l'algèbre des relations (Codd 1972) et l'algèbre des domaines (Merrett 1976) [Codd 1971, 1972 a également établi un calcul des relations]. Une opération exécutée en algèbre des relations est une opération utilisant une ou plusieurs relations comme opérandes et produisant une relation en résultat; l'algèbre des domaines joue le même rôle dans le cas de domaines.

L'algèbre des relations comporte les trois opérateurs de base suivants: réunion, projection et mise en réserve. La réunion est de loin l'instruction la plus puissante existant en algèbre. Essentiellement, elle permet d'associer deux ou plusieurs relations en utilisant comme lien un domaine. A titre d'exemple, la réunion naturelle (que l'on peut assimiler à une opération booléenne AND généralisée) de deux relations ayant un domaine commun donnera pour résultat une relation contenant uniquement les uplets ayant des valeurs identiques dans leur domaine commun. Soit les deux relations ACRO et CORP:

ACRO	ACRONYM	FULL NAME
	ACE	American Council on Education
	ALA	American Library Association
	AACC	American Association of Cereal Chemists

CORP	CORPORATE ACRONYM	CORPORATE CODE
	CIDA	000484
	ACE	004310
	ALA	000062
	ECLA	000573

Si l'on procède à la réunion naturelle de ces deux relations en fonction du domaine ACRONYM, on obtient la relation suivante:

ACRO_CORP	ACRONYM	FULL NAME	CORPORATE CODE
	ACE	American Council on Education	004310
	ALA	American Library Association	000062

Cinq autres réunions ont été définies et peuvent être étudiées dans divers ouvrages (Codd 1972 et Merrett 1976).

La projection est l'opération permettant de générer une nouvelle relation composée d'une ou de plusieurs colonnes d'une table constituant la relation originale. Par exemple, la projection de la relation ACRO_CORP que nous venons de créer relativement au nom complet et au code de l'organisme livrerait la relation suivante:

FULL NAME	CORPORATE CODE
American Council on Education	004310
American Library Association	000062

La mise en réserve (Date 1975) est le nom donné aux opérations relatives à l'insertion et à la suppression d'uplets (ou "notices"). Pour insérer un uplet dans une relation, on a recours à l'union d'ensembles. Celle-ci produit un ensemble contenant tous les éléments des ensembles de départ. A l'inverse, la suppression se fonde sur la différence entre ensembles. Cette opération produit un ensemble contenant tous les éléments du premier ensemble qui ne sont pas contenus dans le second.

Sous sa forme pratique, l'algèbre des domaines a été introduite par Merrett en 1976. Elle permet "de faire des domaines les opérandes et le résultat des opérations courantes d'arithmétique, de logique et d'enchaînement". Dans ce cas, les domaines peuvent alors être virtuels (et le sont habituellement); en d'autres termes, ils n'existent qu'en tant que résultat défini pouvant être obtenu grâce à des instructions de sortie. En algèbre des domaines, les opérateurs peuvent être soit verticaux soit horizontaux. Les opérateurs horizontaux s'appliquent à un ou plusieurs domaines, une ligne à la fois, tandis que les opérateurs verticaux s'appliquent à l'ensemble des lignes, une colonne à la fois. Ainsi, le coût total de tous les livres commandés par une bibliothèque serait le résultat d'une opération verticale; par contre, le nombre de jours que met un fournisseur à remplir une commande serait le résultat d'une opération horizontale (date de réception du livre moins la date de commande).

Selon le modèle relationnel, les "contenants" de base des données sont les relations. Celui qui utilise une base de données (nous l'appellerons l'utilisateur) a accès à ces relations grâce à un sous-modèle de données. Celui-ci permet à l'utilisateur d'appliquer son propre point de vue aux données contenues dans la base de données. Le sous-modèle est actuellement défini par l'intermédiaire d'une définition des données et résulte habituellement d'opérations (s'appliquant à une relation ou à un ensemble de relations) exécutées par les sous-programmes de gestion des bases de données, à la demande d'un programme d'application. Bien que les trois théories de gestion de bases de données comportent une fonction de redéfinition des données, seul le modèle relationnel offre une technique uniforme d'accès aux données de tous les niveaux.

Réalisation pratique

Dans toute mise en place de n'importe quel système de bases de données, il importe de bien définir les données; car si elles sont mal définies, elles ne peuvent être utilisées. Elles doivent avoir une définition structurale et leurs relations avec

les autres données du système doivent être comprises avant qu'on puisse les utiliser. Cette information est essentielle pour les utilisateurs finals de la base de données, pour le système de gestion de celle-ci ainsi que pour la liaison entre les utilisateurs et ce système de gestion. On obtient ainsi trois points de vue différents sur les données: un point de vue extérieur, celui de l'utilisateur, un point de vue intérieur, celui du système d'exploitation, et un point de vue intermédiaire, celui de la mise en liaison des deux premiers. Dans le système du CRDI, une fonction définition de données a été mise en place pour satisfaire à cette exigence.

Relativement au système de bases de données du CRDI, les quatre niveaux d'exploitation suivants ont été définis:

1) utilisateur final: l'utilisateur final du système peut être un chercheur, un bibliothécaire ou un utilisateur occasionnel;

2) gestionnaire du système: cette personne correspond à l'administrateur d'une base de données, dont il est si souvent question dans les ouvrages;

3) programmeur: il s'agit du programmeur d'application, auteur du code assurant la liaison entre l'utilisateur final et le système de gestion de la base de données; et

4) système de gestion de la base de données: à ce niveau, immédiatement supérieur au système d'exploitation, tout accès physique à la base de données est commandé par les sous-programmes.

Voyons d'abord les utilisateurs finals. L'utilisateur définit un sous-modèle de données en fournissant au gestionnaire du système les informations suivantes: 1) le nom que doit porter le sous-modèle; 2) une description des zones du sous-modèle. Cette description comprend la longueur et le nom de chaque zone, une mention indiquant si la zone peut être répétée, le type de chaque zone (numérique ou alphabétique) et une mention indiquant si chaque zone peut être traitée de façon hiérarchique, entre autres: 3) un format d'affichage par défaut et 4) une mention indiquant si ce sous-modèle exige des voies d'accès rapides et les zones permettant de les réaliser.

Le gestionnaire du système traite à nouveau ces informations afin d'obtenir un point de vue intermédiaire sur les données en vue du système. A ce stade, il faut répondre aux questions suivantes: 1) Ces données existent-elles ailleurs? 2) Sont-elles non rémanentes? 3) Des voies d'accès rapides sont-elles justifiées pour le sous-modèle? 4) Les données auxquelles il y a accès doivent-elles être seulement en lecture (consultation) ou aussi en lecture-écriture (consultation et mise à jour)? (Pour formuler ce point de vue intermédiaire, le gestionnaire du système doit procéder avec soin afin d'assurer toujours l'indépendance des données du système.) Le "processeur" de définition des données (voir plus loin) est alors appelé, et le gestionnaire introduit la définition dans le système.

Lorsque ensuite l'utilisateur veut accéder au sous-modèle, le système de gestion des bases de données traite la définition des données, communique le point de vue de l'utilisateur au programme d'application et consulte quant à lui le point de vue interne. (Il faut toujours se rappeler qu'un grand nombre de relations différentes peuvent servir à représenter le sous-modèle de données défini par l'utilisateur. Ces relations peuvent être elles-mêmes réelles ou virtuelles et peuvent ou non représenter un fichier physique unique.) Le point de vue interne est toujours établi au moyen de relations et de domaines. Le Tableau 1 illustre l'acheminement logique entre l'utilisateur final et la base de données physique.

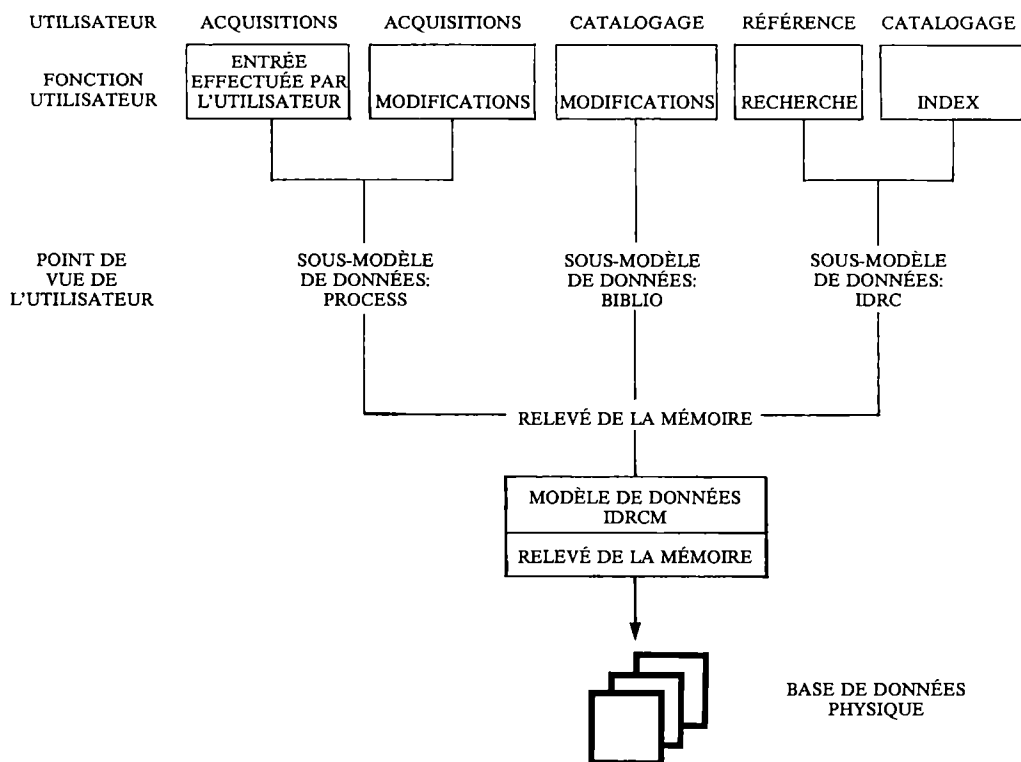


Tableau 1 — Acheminement logique entre l'utilisateur et la base de données physique.

Maintenant que l'on s'est donné un instrument permettant de traiter les diverses formes de données, on peut examiner les fonctions prévues à chaque niveau du traitement de ces données. Il faut alors introduire par ordre d'utilisation les quatre notions suivantes:

- 1) processeur: sous-programme d'application chargé de l'exécution d'une certaine fonction;
- 2) traitement: exécution unique d'un programme par un utilisateur donné à un moment donné;
- 3) chaîne: à la connaissance de l'auteur, ce terme est propre au système HP-3000. L'enchaînement sert à effectuer le "spooling" (l'exécution simultanée d'opérations périphériques et de programmes principaux sur un même système) de travaux (ou de données) par lots, lors de séances de traitement interactif ou de travaux. Le travail ayant fait l'objet du "spooling" est alors planifié par le système d'exploitation et traité de façon indépendante du traitement ou du travail où il a été enchaîné; l'ordinateur passe alors la main au programme d'enchaînement;
- 4) restriction: sélection d'un sous-ensemble d'une relation selon certains critères correspondant habituellement aux valeurs attribuées à des domaines particuliers. Rechercher par exemple tous les articles d'une base de données associés à la Fondation Ford consiste à appliquer à cette base une restriction fondée sur une affiliation.

Les fonctions à chacun de ces niveaux d'exploitation feront maintenant l'objet d'un examen particulier.

Utilisateur final

On a cherché à fournir à l'utilisateur un langage puissant, quoique simple, lui permettant de tirer le meilleur profit du système. Les besoins de l'utilisateur ont été répartis sur les deux catégories suivantes: l'introduction et l'extraction de données. Notre connaissance d'un système automatisé nous a fait prendre conscience de problèmes souvent négligés; à titre d'exemple, des numéros de notices peuvent être un embarras; des notices en double peuvent provoquer bien des ennuis dans le cas d'une collection comptant un aussi grand nombre d'ouvrages reçus en cadeau que d'ouvrages achetés; la standardisation des données peut occasionner des problèmes d'extraction; l'établissement des coûts peut représenter une tâche manuelle fastidieuse; enfin, il est souvent utile de tenir compte par écrit des niveaux bibliographiques. Afin de répondre aux besoins de l'utilisateur, on a donc décidé de mettre en place sept importants "processeurs" d'application.

En introduction de données, on dispose des "processeurs" suivants:

1) INPUT, permettant à l'utilisateur d'introduire de nouvelles notices dans le système. Selon la structure de la donnée à introduire, cette notice peut exiger ou non un numéro d'ordre interne ou ISN (internal sequence number). Lorsque ce numéro est requis, il est généré automatiquement par les sous-programmes de gestion de base de données. Si des niveaux bibliographiques sont nécessaires au point de vue de l'utilisateur, le "processeur" en est alors informé et les opérations requises sont commandées. Par l'intermédiaire de la définition de données, ce "processeur" peut également recevoir l'instruction de vérifier automatiquement les notices en double et de valider le contenu des zones de données au moyen des listes d'autorisations. Si la nouvelle notice est un double, l'utilisateur en est informé et l'ordinateur lui offre la possibilité d'en poursuivre l'introduction ou de procéder à une nouvelle notice; si une zone de données n'est pas validée, l'utilisateur en est également informé; il peut revenir dans la zone après avoir vérifié les listes d'autorisations et trouvé un remplacement valide; il peut également rentrer dans la zone après avoir été envoyé au préalable à un traitement "fils", où il peut créer une entrée autorisant le contenu qu'il vient d'introduire;

2) MODIFY, qui permet à l'utilisateur de modifier les notices inscrites dans le système. Les modifications sont effectuées au niveau des zones et peuvent être l'une des suivantes: CHANGE (modification des données d'une zone), ADD (addition d'une zone), DELETE (suppression d'une zone), TRANSFER (transfert d'une zone à une autre) et REPLACE (remplacement d'une ancienne zone par une nouvelle). Si la notice qu'entend employer l'utilisateur n'est pas accessible par ISN, ou si l'utilisateur ignore l'ISN, celui-ci peut lancer la recherche de la notice en question dans le cadre du "processeur". Si plus d'une notice répond aux spécifications de recherche, l'utilisateur peut sélectionner celle qui convient. Ce "processeur" peut également fonctionner à un niveau global; lorsque cette dernière fonction est demandée, le "processeur" enchaîne en produisant un double de lui-même qui passera à une file d'attente de moindre priorité et permettra ainsi à l'utilisateur d'employer le terminal à d'autres fins, y compris la modification des notices dans le cadre du "processeur" MODIFY.

3) **RELEASE/DELETE**, qui permet à l'utilisateur du système jouissant de la cote de sécurité appropriée d'approuver et de supprimer des notices. La suppression (**DELETE**) se passe d'explication; au moment de la rédaction de cette brochure, de telles suppressions étaient logiques plutôt que physiques. La validation (**RELEASE**), en revanche, n'est pas aussi simple et exige d'être explicitée. Il existe deux types d'approbation d'une notice. Dans le premier type, la notice reçoit un indicateur montrant qu'on ne peut plus la modifier. Cette fonction est particulièrement utile lorsque l'utilisateur du terminal n'est pas l'auteur des bordereaux de traitement ou le créateur de la notice en question; ce type d'approbation signifie que l'article a fait l'objet d'une vérification (qu'il est "propre") et qu'il ne peut pas être modifié accidentellement. Le deuxième type d'approbation supprime le premier, c'est-à-dire qu'il permet une modification ultérieure de la notice.

En extraction de données, on a:

1) **QUERY**, qui permet à l'utilisateur de parcourir les bases de données de façon interactive ou en différé. Il offre aussi bien la fonction de recherche avec les opérateurs booléens que celle en texte libre. Selon les besoins de l'utilisateur, cette extraction peut s'appliquer à l'ensemble ou à une partie de la base de données. Une fonction intéressante de **QUERY** est celle qui permet l'extraction à partir d'un thésaurus multilingue. Elle est rendue possible par l'existence d'un thésaurus véritable et stable. Le **CRDI** emploie le *Macrothésaurus* publié par l'Organisation de coopération et de développement économiques (OCDE) à Paris en 1972, réédité en 1979. Elaboré en partie grâce à des fonds du **CRDI**, ce thésaurus existe dans beaucoup de langues (notamment l'anglais, le français, l'espagnol, l'allemand et le portugais), en associe des descripteurs selon les notions de terme générique (BT), terme spécifique (NT), terme relié (RT), terme "ANY"*, renvoi au terme valide (USE), usage d'un terme valide qui doit être employé à la place du terme interdit (UF) et numéro de facette. Au **CRDI**, pour résumer un document, on emploie les descripteurs du thésaurus comme mots clés intercalés dans un texte. Le thésaurus est disponible au **CRDI** sur ordinateur, en trois langues: français, anglais et espagnol. **QUERY** permet des recherches dans n'importe laquelle des trois langues, comme langue de base et traduit tout descripteur dans les autres langues (en associant automatiquement ces traductions au descripteur original grâce à l'opérateur OR). Cette fonction est particulièrement appréciée par les utilisateurs du Centre, étant donné que beaucoup de documents de notre collection sont rédigés dans une autre langue que le français. **QUERY** permet de plus d'effectuer des recherches dans le thésaurus au moyen de relations structurales à n'importe quel niveau désiré; par exemple, aide au développement (BT) provoque une recherche de ce sujet ainsi que de tout terme d'un sens plus large, dans les trois langues disponibles, à moins que la fonction traduction n'ait été invalidée par l'utilisateur. **QUERY** permet ensuite l'affichage des relations structurales dans le langage de base. Une autre fonction permet de "butiner" (avec ou sans tri) et une autre de rechercher des radicaux;

2) **INDEX**, qui permet à l'utilisateur de produire des sorties dans n'importe quelle séquence, aussi bizarre soit-elle. Ce "processeur" traite la sortie d'une demande de renseignements ou un point de vue de l'utilisateur de manière à produire essentiellement des critères de tri qui serviront à générer des sorties ordonnées à partir d'entrées. Ce "processeur" est l'un des plus puissants du

*Les tables ANY sont employées au cours de la recherche documentaire afin de permettre l'accès à certains blocs d'information.

système et, tout comme MODIFY et QUERY, il peut enchaîner le traitement d'une file d'attente de moindre priorité. Il accepte des spécifications en conversationnel ou à partir d'un fichier redéfini et lance également le travail d'impression nécessaire.

3) COMPUTE, qui permet à l'utilisateur d'effectuer des opérations arithmétiques sur des données et de produire des états à partir des résultats obtenus.

4) PRINT effectue toute l'impression dans le système. Il imprime les sorties d'INDEX et de QUERY de même que toute sortie de la base de données défini par le point de vue de l'utilisateur. Ce "processeur" est d'un emploi très souple; en effet, l'impression peut être faite sur papier spécial ou sur papier ordinaire, sous forme de tableaux ou de colonnes, avec ou sans signes diacritiques, avec ou sans pagination, etc. Le résultat peut être acheminé à un terminal (à imprimante ou à écran cathodique), à une imprimante ligne par ligne ou à tout autre appareil. Les spécifications d'impression peuvent être prédéfinies, puis conservées ou déterminées au moment du traitement dans un échange de type conversationnel dans le cadre du "processeur" PRINT lui-même. Cet échange vise à obtenir de l'utilisateur toutes les informations nécessaires pour traiter le travail d'impression, y compris les spécifications pour déterminer le format des zones. La même fonction permet à l'utilisateur de modifier une spécification existante. Comme tous les autres "processeurs", il peut donner lieu à un enchaînement.

Relativement à l'utilisateur éventuel, les données peuvent être conservées aussi bien en majuscules qu'en minuscules et le codage des signes diacritiques peut être intercalé avec les données (ce codage pouvant être identifié par les "processeurs" concernés). Le jeu de caractères en usage dans le système est l'ASCII à 7 éléments. La longueur maximum de la partie "données" d'une notice a été arbitrairement fixée à 2 048 caractères, ce qui semble être plus que suffisant pour les besoins courants. Le système offre un accès simultané aux bases de données en raison de la fonction de blocage et de déblocage de fichiers que comportent les fichiers du système d'exploitation MPE. Ce blocage s'effectue par granules et au niveau d'une relation (fichier) par opposition au niveau de la notice. (Des études de la littérature actuelle, aussi bien que notre propre expérience, révèlent que cette caractéristique ne porte pas préjudice au rendement.)

Gestionnaire du système

Ce niveau comporte un certain nombre de "processeurs" servant à tenir à jour et à augmenter les bases de données. Ce sont:

1) ISOCONV, qui produit et accepte des bandes de format ISO 2709. C'est celui qui est utilisé pour mettre en place les bases de données reçues d'autres organismes. Il a été conçu de façon aussi générale que possible et comporte des sorties spéciales de traitement, afin que le Centre puisse accepter les données du plus grand nombre possible d'organismes et les utiliser pour des conversions;

2) le "processeur" de définition des données, qui sert à créer, modifier et supprimer les diverses formes de définitions de données (points de vue de l'utilisateur, du système et intermédiaires);

3) RENUMBER, qui permet de remplacer des ISN par d'autres. Ce "processeur" est particulièrement utile à la production de bibliographies imprimées;

4) INVERT, qui sert à effectuer des inversions en lots dans des domaines et à procurer ainsi de nouvelles voies d'accès rapide à la base de données;

5) le récupérateur de positions inutilisées, qui est exécuté périodiquement pour récupérer l'espace libre ou non employé dans les bases de données et sert à générer des doubles des fichiers physiques constituant la base de données;

6) divers "processeurs" d'initialisation.

Le gestionnaire du système dispose également de toutes les fonctions prévues par le fabricant. Heureusement, le HP-3000 en compte beaucoup.

Programmeur

Le programmeur d'application est la personne qui rédige le code servant à exécuter les "processeurs" de l'utilisateur final et du gestionnaire du système. Il dispose d'un ensemble d'appels adressés au système de gestion des bases de données. Les fonctions faisant partie intégrante des sous-programmes de gestion de ces bases permettent au programmeur d'effectuer facilement des opérations comme des demandes de renseignements dans le cadre du "processeur" MODIFY. Les fonctions dont dispose le programmeur pour rédiger des "processeurs" d'application sont des procédures qui permettent de:

- 1) procéder à une analyse syntaxique des dialogues;
- 2) actualiser des domaines virtuels selon les règles de l'algèbre des domaines. Pour le programmeur d'application, cet appel vise le traitement arithmétique;
- 3) associer des contraintes à la base de données; cette procédure peut être utilisée chaque fois que des contraintes sont souhaitables. Ainsi QUERY assure une liaison complexe entre l'utilisateur et le module des contraintes; INPUT l'utilise pour détecter des dédoublements et MODIFY l'utilise pour traiter les demandes de renseignements de l'utilisateur;
- 4) projeter de nouvelles relations à partir de relations existantes;
- 5) effectuer la réunion de relations existantes pour en produire de nouvelles;
- 6) rédiger les notices de l'utilisateur destinées à la base de données;
- 7) lire des notices d'une base de données;
- 8) attester la validité des diverses données (par exemple, lors de l'exécution d'INPUT);
- 9) lire, écrire, supprimer et mettre à jour les zones des notices de l'utilisateur;
- 10) rendre les bases de données disponibles pour le traitement et fournir au programme d'application le point de vue approprié de l'utilisateur en vue du travail à effectuer (par exemple, pour ouvrir la base de données);
- 11) fermer une base de données;
- 12) mettre les notices dans un format pouvant être affiché. Il s'agit là d'une des procédures importantes de PRINT. Elle est également utilisée par QUERY et MODIFY;
- 13) générer au besoin un ISN lors de l'introduction d'une nouvelle notice dans le système;

14) extraire les indicatifs des données. Ces indicatifs peuvent être définis comme des mots, des expressions, des termes du thésaurus, etc. et peuvent être générés sous différentes formes.

Le programmeur dispose également des fonctions prévues par le fabricant, par exemple les procédures SORT/MERGE et TRANSLATE, à partir de l'EBCDIC et de ASCII.

Système de gestion de la base de données

A ce niveau, on examine les procédures servant à mettre en application la théorie des relations. Ces procédures comprennent les quatorze énumérées ci-dessus en plus de trois autres, dont le but est de 1) traiter les indicatifs générés, 2) effectuer le travail concret relatif à l'approbation et à la suppression des notices et 3) organiser les définitions des données de manière à produire le point de vue de l'utilisateur en vue du programme d'application et le point de vue du système en vue du système de gestion de la base de données (ce système est appelé en 10) ci dessus).

Bien que beaucoup de procédures à ce niveau soient enclenchées par des programmes d'application, elles le sont aussi mutuellement pour remplir un grand nombre de fonctions diverses. L'une des raisons est qu'au niveau de la gestion de la base de données, tous les sous-programmes concernent les relations et les domaines, qu'ils soient virtuels ou réels. Afin de produire une notice d'utilisateur permettant de repasser la main au programme d'appel, GETUPLE (la procédure appelée par le programme d'application pour lire une notice) peut avoir à appeler JOIN pour assembler les uplets des relations constituantes. De même, en association avec PROJECT, AUGMENT peut avoir à répartir les notices d'utilisateur en n n -uplets appartenant à différentes relations. A certaines occasions, GETUPLE peut avoir à utiliser PROJECT pour fournir une notice d'utilisateur.

Comme il a été expliqué ci-dessus, il est possible de définir des points de vue de l'utilisateur comprenant des domaines virtuels. Lorsque l'utilisateur tente d'accéder à ce point de vue, la procédure servant à ouvrir les relations détecte qu'il s'agit là d'une relation virtuelle et enclanche la procédure servant à actualiser cette relation pour l'utilisateur. Tout autre traitement de cette relation est identique à celui des autres relations.

Au sujet de l'accès, une remarque s'impose. La théorie des relations ne prévoit pas de voies d'accès non relationnelles à une base de données, mais nous y avons remédié. Des voies d'accès rapide ont été assurées grâce à des fichiers inversés. Cette inversion est effectuée sur les indicatifs extraits des zones de données qui ont été définies pour ce type de traitement par le gestionnaire du système. Le fichier inversé d'un indicatif appartenant à une base de données existante peut être mis à jour aux trois moments suivants: lors de l'écriture d'une notice dans la base de données (dans ce cas, AUGMENT se charge d'appeler les programmes d'indicatifs pour effectuer la mise à jour); lors de la réécriture d'une notice en raison d'une modification quelconque de la zone de données contenant l'indicatif en question (dans ce cas, c'est la procédure de manipulation des zones qui effectue l'appel) ou encore, lors de la validation ou de la suppression d'une notice (dans ce cas, c'est la procédure de la base de données effectuant l'approbation ou la suppression d'une notice qui appelle les programmes). En

réalité, ces trois cas se réduisent aux deux suivants: mise à jour immédiate ou en ligne et mise à jour au moment de la validation seulement. Lorsqu'une notice est supprimée, toute référence à cette notice dans l'ensemble des fichiers inversés est immédiatement supprimée aussi. Pour la réalisation des fichiers inversés, on a eu recours à la technique de mise en mémoire appelée arbres à index B (Knuth 1973); la mise en mémoire des inscriptions et la mise en place de la logique booléenne ont été effectuées grâce à des relevés de mémoire à grande densité en bits.

Remarques au sujet de la mise en place

Le système mis en place n'a exigé aucune modification du logiciel fourni par le fabricant. Bien au contraire, certaines des fonctions mises en œuvre n'ont pu l'être que grâce au logiciel et au matériel de celui-ci. L'ordinateur HP-3000 a l'avantage d'offrir une mémoire à piles superposées. Cela signifie que l'ensemble de notre code est rentrant et récursif. Le logiciel du système est indépendant de l'ordinateur et, par conséquent, notre logiciel l'est également; les opérations peuvent être facilement générées sur simple appel au système d'exploitation; les opérations de l'utilisateur peuvent engendrer, à titre d'opérations, même les "processeurs" du fabricant; tout le traitement au terminal peut être effectué en vertu du logiciel du fabricant (alors que dans le cas d'ISIS, les programmes de traitement des entrées-sorties devaient être rédigés à l'aide de PIOCS, le système de gestion des entrées-sorties physiques d'IBM); les "processeurs" rédigés par le Centre peuvent être traités dans n'importe quel ordre d'une file d'attente, c'est-à-dire qu'ils peuvent être enchaînés, qu'ils peuvent s'enchaîner eux-mêmes ou qu'ils peuvent s'enchaîner entre eux; cette fonction est très importante, étant donné que les "processeurs" habituellement employés en mode interactif peuvent être traités en lots lorsque le genre de travail le justifie; enfin, la sécurité relative à l'accès est assurée par le système d'exploitation. Fait à noter, les fonctions utilisées sont les fonctions de base incorporées au logiciel du système d'exploitation.

Tout le code du système (sous-programmes d'application et de gestion des bases de données) a été rédigé en SPL, le langage de programmation des systèmes du HP-3000. Ce langage de type ALGOL est évolué et constitue également le langage d'assemblage du système. Il permet donc des instructions de structure évoluées comme DO...UNTIL (faire...jusqu'à) et WHILE...DO (faire...tant que) mais également des instructions de manipulation de bits. Il est peu probable que nos efforts en programmation auraient donné d'aussi bons résultats s'il avait fallu rédiger dans un autre langage.

Notre équipe a compté deux personnes pendant douze mois, puis trois personnes pendant sept mois supplémentaires. Au cours de la première période, la conception du système a été réalisée et treize des dix-sept sous-programmes de gestion de bases de données ont été rédigés. Au cours des sept mois suivants, nous avons rédigé trois des quatre sous-programmes de bases de données restants et les "processeurs" de l'utilisateur. L'élaboration du système a été entreprise en juin 1976. Nos premiers utilisateurs ont été desservis en octobre 1977, et par la suite tous les "processeurs" sont entrés en fonction. Il va de soi que tous les "processeurs" du gestionnaire du système sont également opérationnels. Nos réalisations ont fait la preuve qu'une petite équipe dévouée peut assurer l'élaboration aussi bien, voire mieux, qu'un groupe plus important.

La base de données du Centre contient quelque 26 000 références. Le Centre possède également la base de données du BIT, contenant plus de 60 000 références, celle de la FAO, qui est de l'ordre de 40 000 références et deux autres moins importantes contenant au total 10 000 références. Jusqu'à ce que le Centre puisse acquérir une plus grande mémoire de masse, on solutionne le problème de l'accès en direct par un ordonnancement approprié, les bases de données autres que celle du CRDI étant accessibles à certaines heures fixes de la semaine ou sur demande préalable de l'utilisateur.

Bien que notre système de base soit maintenant opérationnel, notre travail n'est pas pour autant achevé. Dans un proche avenir, la procédure relative à l'algèbre des domaines doit faire l'objet d'essais et être mise en œuvre, et nous comptons mettre en place le "processeur" SDI (diffusion sélective de l'information) et terminer la rédaction de la documentation relative au système. Nous espérons également rédiger une procédure utilisant un dérouleur de bande magnétique comme mémoire de masse virtuelle en vue du traitement de bases de données importantes et incorporer à PRINT un programme de liaison de photocomposition.

Il est à souhaiter que ce système gagne le faveur d'établissements ayant besoin d'un système de bases de données bibliographiques, mais qui ne peuvent se procurer un équipement coûteux ni partager un ordinateur de grande puissance. Il est le résultat logique de notre expérience avec ISIS et contribuera, nous l'espérons, à agrandir le réseau commun mis sur pied avec ISIS.

Bibliographie

- CODASYL 1971. CODASYL data base task group report. New York, ACM, avril 1971.
- Codd, E. F. 1970. A relational model of data for large shared data banks. CACM, 13(6), 377-397.
1971. A data base sublanguage founded on the relational calculus. Proceedings of 1971 ACM-SIGFIDET Workshop. New York, ACM.
1972. Relational completeness of data base sublanguages. Rustin, R., rédacteur principal, Database systems, Englewood Cliffs, N.J., Prentice Hall.
- Date, C.J. 1975. An introduction to database systems. Reading, MA, Addison-Wesley, Publ. Co. Inc.
- Eswaran, K. P., Gray, J. N., Lorie, R. A., et Traiger, I. L. 1976. The notions of consistency and predicate locks in a database system. CACM, 19(11), 624-633.
- Knuth, D. E. 1973. The art of computer programming, volume 3: sorting and searching, Englewood Cliffs, N.J., Addison-Wesley, Publ. Co. Inc.
- Merrett, T. H. 1976. MRDS — an algebraic relational database system. Proceedings of the Canadian Computer Conference, Montréal, mai 1976, 102-124.
- Organisation internationale de normalisation. 1973. ISO-2709 Documentation — Disposition des données sur bande magnétique pour l'échange d'informations bibliographiques.
- Organisation pour la coopération et le développement économiques. 1972. Macrothésaurus: Liste des principaux descripteurs relatifs au développement économique et social. OCDE, Paris.
1979. Macrothésaurus pour le traitement de l'information relative au développement économique et social. Nouvelle édition française. OCDE, Paris.
- Ries, D. R. et Stonebraker, M. 1977. Effects of locking granularity in a database management system. ACM Transactions on Database Systems, 2(3), 233-246.

